

Developing the Program Evaluation Utility Standards: Scholarly Foundations and Collaborative Processes

Donald B. Yarbrough
University of Iowa

Abstract: *Developing the third edition of the program evaluation utility standards required multilevel collaborations among task force members, members of the Joint Committee on Standards for Educational Evaluation, sponsoring organizations, and hundreds of involved stakeholders. The scholarship on evaluation use, influence, and collaboration was foundational for the utility standards and materials accompanying them and equally important for informing the processes guiding utility standards development. This article emphasizes the foundational role of this recent scholarship and the roles played by all who collaborated in planning and implementing the utility standards development processes.*

Keywords: *evaluation, Joint Committee on Standards for Educational Evaluation (JCSEE), standards, utility*

Résumé : *Il a fallu une collaboration à toutes sortes de niveaux entre les membres du groupe de travail, les membres du Comité mixte sur les normes d'évaluation en éducation, les organisations commanditaires et des centaines d'intervenants pour arriver à la troisième édition des normes d'utilité en évaluation de programme. Les normes d'utilité et les documents connexes ont leur fondement dans les recherches sur l'utilisation, l'influence de l'évaluation, et la collaboration, recherches qui ont eu un impact important sur les processus qui ont guidé l'établissement des normes d'utilité. Le présent article met l'accent sur le rôle fondamental de ces recherches récentes et sur les rôles joués par toutes les personnes qui ont contribué à la planification et à la mise en œuvre des processus d'élaboration des normes d'utilité.*

Mots clés : *évaluation, Comité mixte sur les normes d'évaluation en éducation, JCSEE, normes, utilité*

The purpose of this article is to describe the processes that resulted in the third edition of the *Program Evaluation Standards* (Yarbrough, Shulha, Hopson, & Caruthers, 2011) and to document how these processes and the resulting standards were informed by scholarship on evaluation use, utility, influence, and collaboration. In honour of Lyn Shulha's lifetime achievements, this article emphasizes

Corresponding Author: Donald B. Yarbrough, University of Iowa, Iowa City, Iowa, 52246, USA, d-yarbrough@uiowa.edu

how her scholarly contributions along with those of others informed the content expressed in the third edition utility standards and also led to refined collaborative methods in the standards development processes. In particular, the scholarship on use and collaboration in evaluation (e.g., [Shulha & Cousins, 1997](#); cf. [Shulha, Whitmore, Cousins, Gilbert, & al Hudib, 2016](#)) led to more open approaches to collaboration in the development of the third edition standards.

The third edition of the program evaluation standards was intended to address three major purposes. First, the standards are expected to address current evaluators' and evaluation users' needs. To facilitate this purpose, the procedures of the American National Standards Institute (ANSI) and Joint Committee on Standards for Educational Evaluation (JCSEE) required that all approved standards undergo a major revision process that included current stakeholders at least every 10 years. Because the second edition was completed in 1994, the standards were out of date without revision by 2004 (JCSEE, 1981, 1988, 1994; [Yarbrough, Shulha, & Caruthers, 2004](#); cf. [Yarbrough et al., 2011](#)).

Second, the standards are intended to fully reflect and align with current and past evaluation theoretical and practice scholarship. When aligned with scholarship, the standards can then serve as efficient scaffolding for knowledge and values related to evaluation and help facilitate better evaluation practice (cf. [Yarbrough et al., 2004](#)). Much of this article is directed to documenting how the scholarship on evaluation use, utility, and collaboration provided a foundation for the utility standards development processes and the products that resulted (cf. [Yarbrough et al., 2011](#)).

The third intended purpose for this edition, also informed by scholarship, was to clarify dimensions and subdimensions of quality in evaluation processes and products as currently understood. In response to stakeholder input and reflecting current practice and theoretical scholarship, the third edition organized the standards into five rather than four dimensions of quality as in previous editions. The fifth dimension of quality, evaluation accountability, was added to utility, feasibility, propriety, and accuracy to organize the revised standards into sets supporting these five specific dimensions of quality in evaluation.

The emphasis on utility in this article is not meant to detract from the importance of the other dimensions of quality in evaluation and their supporting domains of scholarship. However, reviewing the scholarship supporting all standards is beyond the scope of this article. Additional articles discussing scholarly foundations of the other four foundational dimensions of evaluation quality (*feasibility, propriety, accuracy, and accountability*) are in preparation. Some areas of scholarship that were paramount for all dimensions—for example, the scholarship on cultural responsiveness and cultural competence—are not included in this article because of space limitations and are being addressed in other work in progress.

NATURE OF THE PROGRAM EVALUATION STANDARDS

The third edition program evaluation standards are different from other kinds of standards in several important ways. First, they do not specify how evaluation

processes will be implemented exactly for each specific context. They do not address exact procedures or product specifications (e.g., technical specifications for the universal serial bus or USB port). Rather, they require reflection and situated application. Second, these are not content standards that mandate the domains of knowledge for curriculum and instruction or assessments of competence in program evaluation (Yarbrough et al., 2011). Instead, the program evaluation standards identify the specific domains, dimensions, and subdimensions of quality in evaluation processes and products. The supporting materials in the standards book (e.g., implementation guidelines, hazards) identify multiple tasks that application of the standards can help address, often including relatively ill-defined problems and problem spaces (Newell & Simon, 1972). To encourage reflective practice and application, these materials provide heuristics that guide and inspire, rather than algorithms that are to be implemented (see Yarbrough et al., 2011, for specific examples).

Developing the standards themselves entailed complex problem-solving tasks, varying in their difficulty of definition and contextualization (cf. Newell & Simon, 1972; Simon 1973). For example, one step in building the scholarly foundations for the utility standards was to identify the domains of research and practice scholarship that should be reviewed and then to craft ways for that knowledge to inform the standards and ancillary materials (e.g., case scenarios). Both direct and indirect processes helped identify and apply this scholarship to the standards. At the end of each of the 30 standards chapters, 5 to 10 key pieces of theoretical and applied scholarship document the contents of that chapter and provide starting places for standards users who want to read more. In addition, thorough compilations of references at the back of the third edition book include additional scholarly sources supporting the content related to the specific sets of standards and their application.

BACKGROUND AND SPECIFIC TASKS IN DEVELOPING THE PROGRAM EVALUATION STANDARDS

The procedures used to develop the first two editions of the program evaluation standards are described in published scholarship (Sanders, 1994; Yarbrough et al., 2004). For the first two editions of the standards, the members of the JCSEE themselves met and wrote the full standards chapters, sometimes supported by drafts from a panel of writers. The format for chapters and the grammatical form of the standard statements themselves were voted on and approved in advance by the JCSEE membership, with guidance from the JCSEE chairperson (for the first edition, Daniel L. Stufflebeam, chair from 1975 to 1988, and for the second edition, James R. Sanders, chair from 1988 to 1998). Feedback for revisions to the original drafts came from national and international reviewers, from national hearings, and from field trials, in that order, in keeping with JCSEE procedures. Feedback from external sources was integrated into the five required drafts of the standards by small groups of JCSEE members and by the JCSEE committee chair. Turnover in JCSEE membership often resulted in different approaches to

drafting and redrafting, with much of the finalization resting on the chairs and a few of the members.

By the time of the development of the third edition, important factors necessitated refinement of the procedures for finalizing the standards and the ancillary materials. First of all, the JCSEE membership no longer met for extended durations of time (a week or more) during which they could devote their energies to drafting and redrafting standards. In addition, by 2004 the JCSEE had three different books of standards to manage (the *Program Evaluation Standards*, *Personnel Evaluation Standards*, and *Student Evaluation Standards*, <http://www.jcsee.org>). JCSEE members could no longer devote all their energies to developing and revising new editions of any one book.

In 2003, the JCSEE officially voted to begin planning for the development of the third edition and to authorize the then chair of the JCSEE, Arlen Gullickson (1998–2008), to appoint an acting chair of the Program Evaluation Standards Development Task Force (JCSEE, 2003). This step was in keeping with the task force model for standards development, which had been partially implemented with the personnel evaluation standards (JCSEE, 2009). Following the task force model, JCSEE members did not draft the standards names, statements, or chapters; rather a small group of volunteers (the task force), under JCSEE oversight and approval processes, drafted multiple successive approximations up to and including the final versions of the standards names and statements and finally the full book. Submitted drafts became final only with the approval of the JCSEE membership. Guidelines approved by the JCSEE required the task force chair to be a JCSEE member (Yarbrough, 2005a, 2007b).

By 2004, the JCSEE had formally approved the Program Evaluation Standards Development Task Force to include Flora Caruthers of the National Legislative Program Evaluation Society; Rodney Hopson, a nonsponsored JCSEE associate member; Lyn Shulha of the Canadian Society for the Study of Education; and Don Yarbrough of the National Council on Measurement in Education, also task force chair (JCSEE, 2004). All members were volunteers, approved formally by vote of the JCSEE. Membership on the task force stayed open for the entire duration of the development, but no additional people volunteered or were added. The task force was diverse with regard to gender, type of professional evaluation work, sponsoring organizations, background and ethnicity, geography, and to some extent age, although all of us were at least midcareer. In addition to being sponsored by different organizations, we were active members of other North American professional organizations as well. For example, we were all long-term members of the American Evaluation Association (AEA). In addition, we remained active throughout the approval processes including final publication and dissemination of the standard names and statements (cf. JCSEE, 2008). The work was finished when the third edition book with its introductions, explanations, case scenario applications, scholarly foundations and documentation, and ancillary materials, was published in 2011, approximately one year after the standard names and statements were approved in final form and posted online (<http://www.jcsee.org>; JCSEE, 2016; cf. Yarbrough et al., 2011).

The formal tasks that guided the development of the third edition of the program evaluation standards were as follows:

- reviewing previous editions of the standards and their uses to determine how the third edition might be improved;
- assessing current needs related to standards across broad groups of stakeholders with regard to aspects of the standards to be retained and improved;
- preparing revisions and new drafts of standard names, statements, and supporting and ancillary materials;
- arranging monitoring and oversight of the procedural steps by an external validation panel that held the developmental processes accountable to the specified JCSEE procedures and the ANSI *Essential Requirements* (cf. ANSI, 2010, 2017; the members of the Program Evaluation Standards Validation Panel and their work are described in their report and summarized in the standards book, [Alkin et al., 2008](#); [Yarbrough et al., 2011](#));
- providing regular updates and process monitoring to the JCSEE for annual review of progress (e.g., [Yarbrough, 2005a, 2007b](#));
- finding and engaging national and international reviewers to provide diverse feedback on near-final drafts of the revised standards;
- conducting no fewer than five formal hearings at national conferences of member organizations and web-based hearings open to all stakeholders to collect information about the quality and value of the drafted revisions;
- revising the multiple draft manuscripts, directly and explicitly responding to stakeholders' and JCSEE members' suggestions and comments;
- finding and engaging large numbers of stakeholders in independent field trials of the revised draft standards and materials and responding to their suggestions;
- implementing final monitoring and comment periods in accordance with the latest ANSI procedures as part of the finalization of the new edition before dissemination;
- scheduling periodic and final JCSEE votes leading to the final version of the standards;
- disseminating and continuing research on standards use and quality;
- identifying and preparing for new updates requiring varying degrees of revisions during the next 10-year cycle; and
- facilitating continued JCSEE use, oversight, and input to create the strongest possibilities for positive influences on evaluators and evaluation users.

One of the first organizing decisions was to divide up literature reviews such that two task force members were primarily responsible for reviewing the scholarship on each major dimension of quality. Shulha and Hopson took the lead on *utility* (including factors affecting use, usefulness, and influence). Caruthers and Yarbrough took the lead on *feasibility* (including factors affecting evaluation

project management, costs-benefits, and evaluability). Hopson and Shulha took the lead on *propriety* (including human rights, evaluation ethicality, legality, morality, and professionalism). Yarbrough and Caruthers took the lead on *accuracy* (including aspects related to descriptive and evaluative attributions based on the full gamut of naturalistic to controlled designs and qualitative and quantitative information types, including mixed methods and designs). Lastly, Yarbrough, in collaboration with Shulha, Hopson, and Caruthers, took the lead on *evaluation accountability* (accountability methodologically supported by internal and external summative and formative meta-evaluation). In addition to the usual research review methods, task force members sought out numerous other scholars and stakeholders with knowledge about these domains to recommend scholarship for review and key topics to be addressed.

Implementation of these tasks took place at different times and at different speeds, with some tasks continuing before, during, and after the development period (e.g., summary and application of scholarship, research on standards). In general, the development process began in 2003 with the JCSEE authorization to undertake the major revision and ended in 2010 after JCSEE approval of the standards in their final form and the finalization of the book for publication. Fourteen members, 12 in attendance and 2 via e-mail, voted for approval, and one member not in attendance voted via e-mail to abstain (ANSI audit files, available at <http://www.jcsee.org> by request).

The scholarship typically associated with evaluation utility, especially related to collaboration as a method to increase evaluation utility, was especially important for all the standards development processes. Not only did this scholarship provide a basis for developing the content of the utility standards and ancillary materials, but it also informed the way that task force members collaborated with one another and planned interactions and collaborations with other stakeholders in the development processes.

DEVELOPMENT OF A THEORY OF STANDARDS UTILITY, USE, AND INFLUENCE

The scholarship on evaluation use, utility, and influence (cf. Cousins & Shulha, 2006; Shulha & Cousins, 1997) was very important for the development of the utility standards in two specific ways. First, that scholarship informed the content to be reflected by the utility standards and the ancillary materials. The scholarship was the foundation for the eight standard names and standard statements (U1 through U8) as drafted by the task force and approved by the JCSEE (JCSEE, 2016). The scholarship was also foundational for all utility components in the book, including the utility overview (pp. 3–9), the utility case scenario (pp. 9–14), and the utility references (pp. 294–304). In addition, each of the eight separate chapters describing the utility standards contained its own unique scholarship-based rationale and clarification (pp. 15–17, 23–24, 29–31, 37–39, 45–47, 51–53, 57–58, 65–67). The chapters also had sections on implementing the standard,

possible hazards, and applications of that standard to the utility case scenario—all scholarship-based. Each individual standard chapter closed with documentation (pp. 22, 28, 35, 43, 50, 55–56, 62–63, 70). Documentation included 5 to 10 references that were specific to each standard, carefully selected for value and clarity by the task force members and collaborators, and then further vetted by reviewers, national hearing participants, and field testers.

From the perspective of the American National Standards Institute (ANSI), an approving body for these standards along with the JCSEE, all standards must be rooted in technical literature (ANSI, 2010). That is one purpose of the documentation sections. A second major purpose was to encourage deep processing of the standards through the ancillary materials, with the documentation section representing initial scholarship that users of the standards could review to enrich their own deep knowledge of the constructs described and operationalized in that particular standard's chapter. Especially for new users, such self-study and professional development would provide the basis for the reflective applications of the standards to the real world program evaluation situations that reflective practitioners would encounter.

The second way in which the scholarship on use, utility, and influence played a role in the development of the program evaluation standards was equally important. For the development of the standards to be maximally useful to stakeholders, the standards development *processes and outputs* had to be useful to stakeholders. Stakeholder collaboration was a key value in standards development (cf. Alkin et al., 2008; Yarbrough, 2005a), just as it was also an essential requirement for approval by ANSI (2010; 2017). Collaboration in the development process was a prerequisite for the production of highly valued and useful standards accompanied by suggested procedures and processes that engaged and provided value to future users.

For purposes of emphasis, Table 1 first provides an overview of some specific types of appropriate uses that the development processes would support, aligned with types of uses of evaluation processes more generally. It also provides an overview of types of uses that can be made of the utility standards and ancillary products specifically and of evaluation products (especially findings and reports) more generally.

Numerous possibilities for addressing and preventing misuse of the standards and the standards development processes also emerged from the scholarship on evaluation misuse and helped guide the collaborations (cf. Cousins, 2004; Cousins & Shulha, 2006; Yarbrough et al., 2011). The scholarship on evaluation influence (e.g., Henry, 2003; Mark & Henry, 2004; Kirkhart, 2000) was also foundationally important for the utility standards (e.g., standard “U8 Concern for Consequences and Influence”) and for aspirations with regard to long-term influences of the standards development processes and subsequent standards dissemination. Because these long-term influences are important, more empirical studies investigating influences of the standards development processes and the standards themselves are needed.

Table 1. Process and Product Uses for Evaluation and Utility Standards Development

Type of Use	Example	
	Evaluation	Utility Standards Development
Process		
Instrumental (action)	Stakeholder and groups who participate in the evaluation processes (e.g., program description processes or focus group or other interviews) may learn about program weaknesses, leading immediately to improvements in the program before any evaluation results.	Stakeholders conducting a field trial may experience skill growth, leading to improvements in unrelated evaluation designs or planning documents that they are currently working on.
Conceptual (understanding)	Program staff collaborating on an evaluation may experience growth in knowledge about the potential value of evaluation for the organization.	Stakeholders who participate in the review of the draft utility standards may grow in evaluative thinking capacities.
Product		
Instrumental (action)	Stakeholders use the evaluation report to improve the program and also as a design format for next year's evaluation plan.	Practicing evaluators who are conducting a field trial use the near-final draft standards and later use the approved standards to improve their evaluation proposals to funding agencies.
Conceptual (understanding)	Program staff who are reading the final report experience growth in knowledge about the potential value of evaluation for their organization.	Stakeholders who read the utility standards (and/or the standards book) experience deeper understanding of evaluation knowledge and strengthened evaluative thinking capacity.

Note. Evaluation and standards development processes refer to any activities as experienced by those involved; evaluation products include reports and other documents; the utility standards development products include the standards themselves and the standards book.

Types of Uses and Specific Uses of the Utility Standards

As summarized in [Table 1](#), the standards—just like evaluations themselves—can be used for both instrumental purposes and conceptual purposes (cf. [Alkin & Taut, 2003](#); [Shulha & Cousins, 1997](#)). For example, they can be highly valuable in helping reflective users implement evaluation processes and products for improvement or accountability purposes (instrumental uses). The meta-evaluative information that

applications of the standards provides (e.g., via “desk audits,” [Yarbrough et al., 2011](#), p. 275) also serves instrumental use. Similarly, many users who apply the standards discover in the process of reviewing, considering, and applying them a more profound understanding of evaluation quality, which illustrates a conceptual use. Applying the standards reflectively across multiple evaluation contexts can lead to improvement in future evaluation designs and implementations and more sophisticated evaluative thinking, a cascade of conceptual and instrumental uses, reflecting the complexity and interactivity of use (cf. [Shulha & Cousins, 1997](#)).

Theories of change related to program evaluation standards development and use were an important part of planning (cf. [Weiss, 1998](#)). Working backwards from the hoped-for time in the future when the new standards would be approved and welcomed by evaluators and evaluation users, we asked ourselves what might be some important uses made of the new standards and how we could facilitate processes to result in the most effective standards for these uses. We divided the possible uses into various organizations and categories. At the most general level, we started with four general categories of possible uses. We wanted the standards to address multiple related areas of needs, such as the need for

- improvements in specific evaluation work in the real world, in real time;
- knowledge and skill growth acquired through teaching and learning about evaluation, including evaluator training in formal and informal settings such as self-study, graduate and professional programs, and professional development more generally;
- increased evaluation capacity, especially related to the development of evaluation users and evaluative thinking capacities, both in specific individual users and in organizations more generally; and
- increased evaluation accountability, both by making the case for the cost-benefit of conducting more program evaluations and by increasing the quality of evaluations as they take place.

A full discussion of all the needs to be addressed by standards development and by the finalized standards and ancillary materials is beyond the capacity of this article. The focus on needs, uses, and utility of the standards is illustrated in detail in the case scenarios and scenario applications across individual standards chapters in the standards book. Specific needs that the standards can address are also detailed throughout the book, especially in the “Introduction,” “Applying the Standards,” the “Functional Table of Standards: Applying the Standards to Some Common Evaluation Tasks,” and the individual standards chapters ([Yarbrough et al., 2011](#)).

Scholarly Foundations for Collaboration in the Development of the Program Evaluation Utility Standards

Foundational scholarship on collaboration was also critical to the development of the utility standards specifically and other standards more generally (cf. [Thomson,](#)

Perry, & Miller, 2007; Wood and Gray, 1991). Recent reviews related to collaborative approaches to evaluation further elaborate these ideas (cf. Cousins, Shulha, Whitmore, Al Hudib, & Gilbert, 2016; Cousins, Whitmore, & Shulha, 2013; Shulha, Whitmore, Cousins, Gilbert, & al Hudib, 2016). Just as with the utility scholarship, research and theory on collaboration contributed to the foundational content for the utility standards and served to enlighten the standards development process.

With regard to the standards development processes, a major concern was how to build the most useful processes to facilitate collaboration among members of the task force, JCSEE members, members of the sponsoring organizations, and all other involved stakeholders. We wanted our procedures to be based on scholarly principles for effective collaboration as experienced by participating stakeholders in these processes, including evaluators and users of evaluation who were contributing but not directly responsible for the final standards products. Just as with scholarship on use, utility, and influence, we focused on using the scholarship on collaboration to ground the key constructs in the utility standards so that the standards could encourage and facilitate highly effective collaborations in program evaluation design, implementation, and use.

Collaboration and participatory approaches to human enterprises, of course, are not unique to evaluation or standards development. Scholarship in the past decades on collaboration has focused on numerous areas of human activity in which collaboration is essential to problem solving. Some examples are business and organizational behaviour; various domains of scientific research and policy formulation; and public health, educational, and other social programming (cf. Bennett & Gadlin, 2012; Thomson et al., 2007; Wood & Gray, 1991).

The key ideas from collaboration scholarship that undergird the utility standards are also well reflected in the burgeoning research and practice literature describing the complex ways in which evaluators and other stakeholders work together to achieve beneficial and useful evaluations (Shulha & Cousins, 1997). Earlier work on evaluation collaboration often emphasized the value of stakeholder participation in the evaluation and its impact on utility and use through the joint identification of evaluation context factors and stakeholders' needs that evaluation might address (cf. Cousins & Earl, 1992; Cousins & Whitmore, 1998). Scholars also began to argue for *collaboration* defined as a more active state of sharing of responsibilities, leadership, and decision-making (O'Sullivan, 2004). To take on the issue of power imbalances more directly, some scholars sought to change the power dynamics among evaluators and different groups of stakeholders to advantage otherwise less powerful groups and individuals (cf. Fetterman, 2001). Most recently, in the time after the publication of the utility standards, scholarship has also extracted evidence-based principles to guide collaborative evaluations (Shulha et al., 2016).

During the development of the third edition of the program evaluation utility standards, we task force members discussed in detail what evaluators should do specifically with regard to creating wide-ranging collaborative processes that enhanced evaluation utility, use, and influence. Gleaning the best advice from many sources, Shulha and others reviewed participatory, collaborative, democratic (House & Howe, 2003), and empowerment approaches. Shulha then drafted several key directives using collaborative processes to enhance utility. These resonate well with the eight principles extracted from more than 300 practicing evaluators several years later (Shulha et al., 2016).

Selected process goals for increasing evaluation utility (one for each utility standard) through collaborations among stakeholders are presented in Table 2. To illustrate alignment and congruence, these goals and the specific eight utility standards are then crosswalked with the eight principles to guide collaborative approaches to evaluations (Shulha et al., 2016).

As the crosswalk in Table 2 illustrates, these evidence-based principles to guide collaborative approaches are connected to various process goals that are explicitly developed in the utility standards. While individual process goals primarily relate to different constellations of principles, all principles for collaborative approaches are implicit in at least one or more of the process goals. For example, all the process goals are related to *motivation for collaboration*, because effective collaboration is required for each of these process goals to be achieved. Similarly, *meaningful relationships* and *appropriate participatory processes* among collaborators are required for attaining all of the process goals. *Developing a shared understanding of the program* is directly related to meeting three of the process goals mentioned in Table 2; *resource availability* is related directly to one process goal (but is also very important for other dimensions of quality, for example, feasibility). *Follow through* and *evaluation quality* seem to be directly important for six of these process goals, and *evaluative thinking* seems to support five of them. These eight process goals that are explicit in the utility standards and ancillary materials are but a small sample of explicit and implicit process goals supporting utility that can be identified from a reflective reading of the standards and ancillary materials (cf. Yarbrough, 2016).

The principles to guide collaborative approaches can also be related to other process goals to increase utility (as well as other dimensions of quality—accuracy, propriety, feasibility and evaluation accountability). Because it is for demonstration purposes only, this crosswalk might not exactly match other evaluators' or evaluation users' attempts to relate process goals associated with the utility standards to these principles. For greatest benefit, individual reflective practitioners should undertake specific crosswalks for targeted evaluation work to achieve a greater understanding of how implementing these principles can facilitate process goals that are essential for the utility of a specific evaluation in its context.

Table 2. Dimensions of Quality in Implemented Evaluations: Eight Selected Evaluation Process Goals to Increase Utility, Cross-walked with Principles to Guide Collaborative Evaluation Approaches

Utility Standards		Related Principle to Guide Collaborative Approaches						
Selected Process Goals in Evaluations that Further Utility and Use	Related Standard	Mot	Rel	Und	Part	Res	Qual	Think Follow
Maintain or increase evaluator credibility over the course of the evaluation by (a) creating a climate of perceived legitimacy, trustworthiness, and competence; (b) demonstrating professionalism; and (c) collaborating effectively and efficiently	U1 Evaluator Credibility: Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.	X	X		X			X
Identify and attend to persons involved in the program and its context whose needs could be addressed by the evaluation	U2 Identified Stakeholders: Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation (Also U1, U3)	X	X	X	X		X	X
Identify evaluation purposes and needs to be addressed based on involvement, negotiations, and renegotiations with adequate numbers and groups of stakeholders	U3 Negotiated Purposes: Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation (Also U1, U2)	X	X		X		X	X
Clarify and specify the diverse individual and group values underlying purposes, judgments, and program and evaluation processes and methodologies	U4 Explicit Values: Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments (Also U1, U2, U3)	X	X	X	X		X	X

(Continued)

Table 2. Continued

Utility Standards		Related Principle to Guide Collaborative Approaches						
Selected Process Goals in Evaluations that Further Utility and Use	Related Standard	Mot	Rel	Und	Part	Res	Qual	Think Follow
Attend to the continuing and emerging needs of multiple evaluation stakeholders and negotiate and renegotiate the intended evaluation purposes to increase utility	U5 Relevant Information: Evaluation information should serve the identified and emergent needs of stakeholders (Also U1, U2, U3, U4)	X	X		X		X	X
Create meaningful evaluation processes and products from the perspectives of stakeholders and encourage stakeholders to reflect on their understandings of and judgments about the program, as well as their roles in the evaluation	U6 Meaningful Processes and Products: Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviours (Also U1, U2, U3, U4, U5)	X	X	X	X		X	X

Develop and implement a communication plan that addresses the full timeline for communication among stakeholders in their various cultures, indicating how (a) stakeholders would be identified and included in discussions of key evaluation processes, products, and available resources; (b) data and information would be requested from stakeholders; (c) information about evaluation participation and collaboration would be shared; and (d) stakeholders could access, interact with, and engage in the preparation and use of formal and informal reports and other evaluation products	U7 Timely and Appropriate Communication and Reporting (Also U1, U2, U3, U4, U5, U6)	X	X	X	X	X	X	X	X
Promote responsible, accurate, adaptive, and reflective uses of information, findings, and conclusions in collaboration with key stakeholders in their various cultural contexts	U8 Concern for Consequences and Influence (Also U1, U2, U3, U4, U5, U6, U7)	X	X	X	X	X	X	X	X

Note. Columns 3–10: Mot = “Clarify motivation for collaboration”; Rel = “Foster meaningful relationships”; Und = “Develop a shared understanding of the program”; Part = “Promote appropriate participatory processes”; Res = “Monitor and respond to resource availability”; Qual = “Monitor evaluation progress and quality”; Think = “Promote evaluative thinking”; Follow = “Follow through to realize use.” These names of principles to guide collaborative evaluation approaches (Columns 3–10) are adapted from [Shulha et al. \(2016\)](#). (See figure on p. 194 and discussion starting on p. 198 in [Shulha et al. 2016](#).) Column 1: The selected process goals related to evaluations that further utility and use in Column 1 are adapted from “Part 1. Utility” ([Yarbrough et al., 2011](#), pp. 3–70). Column 2: Standard names and statements in Column 2 are from the same volume and also available at the JCSEE website (2016).

EXAMPLES OF COLLABORATIVE PROCESSES DURING DEVELOPMENT OF THE THIRD EDITION STANDARDS

At the same time the collaborations were taking place, ongoing review of scholarship was leading to new insights about how to make the collaborations most beneficial. All of us on the task force had been engaged in collaborations for many years. However, Lyn Shulha helped incorporate the best of what we knew and were coming to know from the collaboration scholarship into the standards development processes.

All 16 JCSEE members and hundreds of other volunteers not only provided input and oversight but also assisted with recruiting reviewers, national hearing participants, and field testers. Volunteers from member organizations provided numerous contributions, including a cultural reading of the previous program evaluation standards for cultural strengths and areas for improvement ([American Evaluation Association Diversity Committee Cultural Reading Task Force, 2004](#)). All in all, more than 400 people who participated in the development processes were recognized by name in the third edition. Although no demographic data were collected and analyzed, some additional speculation about these collaborators' backgrounds is possible based on names and organizations provided. For example, with regard to the approximately 100 national and international reviewers, 50 were affiliated with an institution of higher education. Approximately 20 were affiliated with a governmental office or agency, and approximately the same number with a private firm, nonprofit, or foundation. Five reported a primary affiliation with a school or school district. Five did not provide information that allowed easy classification. Of the approximately 240 national hearing participants who provided information (across five national sites and one web-based portal), approximately 85 were from governmental offices or agencies, 90 were from institutions of higher education, 45 were affiliated with firms, nonprofits, or foundations, 15 reported a primary affiliation with a school or school system, and six did not indicate a professional affiliation. Classification is somewhat imprecise because some reviewers had multiple organizational roles and experiences and were perhaps expressing only one of multiple affiliations when providing information. Estimating the number of responses from people primarily located in countries other than the United States was also difficult. Approximately one third of responses indicated a primary affiliation outside the US, with Canada being the primary source of contributions, and other locations such as Turkey, New Zealand, and France mentioned only infrequently (cf. [Yarbrough et al., 2011](#)).

The contributions of individuals listed in this group of 400 ranged from one-time suggestions in a few sentences or paragraphs to ongoing, multiyear comments and discussions about most of the facets of the standards development and the resulting products. Collaborating volunteers helped with all aspects of standards development, from reviewing scholarship, to serving as invited discussants for task force conference presentations, to line-by-line responses to drafts of various materials. We actively solicited collaborators from diverse backgrounds;

all volunteers were encouraged to contribute and participate in the processes. An organizing schema for the collaborations would include the following categories:

- Rich, dynamic, reciprocal collaborations among task force members, in response to self-generated topics and issues as well as input from other collaborators
- Bidirectional discussions based on JCSEE committee members' oversight, insights, suggestions and guidance, often in response to products provided to them by the task force
- Individual interactions with reviewers, including written reviews, in-depth discussions, and exchanges of notes
- Individual and group interactions with national hearing participants, stakeholders attending presentations at national conferences, and other interested individuals who contacted task force or JCSEE committee members based on their interest in the standards
- Individual interactions with volunteers who wanted to discuss their needs related to program evaluation standards and to provide their suggestions about content and format of the standards
- Group interactions with representatives of the Validation Panel, the JCSEE, and task force members.

In implementing the program evaluation standards development processes, we strived for horizontal processes that included stakeholders from a wide diversity of backgrounds and cultures. All were in agreement that a major area in need of attention during the standards development process was cultural awareness, cultural competence, and cultural responsiveness in the standards and standards development processes (cf. [Botcheva, Shih, & Huffman, 2009](#); [Hood, 2005](#); [Kirkhart, 2005](#)). The development process was also unfolding when multiple professional organizations, including those sponsoring members of the JCSEE (e.g., AEA, AERA, APA), were developing or revising policies on cultural competence and cultural responsiveness. Currently, whether the third edition utility standards can be used effectively in other languages and cultures is under empirical evaluation (cf. [Yarbrough et al., 2016](#)).

At the beginning of the third edition standards development process, one of the first tasks was to conduct a national survey of stakeholders' identified needs for changes in the format, structure, and content of the standards. Approximately 340 members of JCSEE member organizations responded ([Yarbrough, 2005b](#)). This led to a second survey further investigating possible preferred standard statement formats ([Yarbrough, 2007a](#)). Whereas previous standards development efforts had relied on prespecified grammatical forms and wording, the third edition approach opened up input on fundamental structural features of the standard statements and chapter organization. One result was that the format used to draft the standard statements for approval by the JCSEE was simpler and more flexibly responsive to content suggestions than the standardized formats used in previous

editions (cf. [Yarbrough et al., 2011](#)). In addition, suggestions from stakeholders inspired changes in the organization, continuity, and length of case scenarios.

We task force members met in person as a group more than 20 times, usually for two or three days. We also convened frequent, multihour conference calls and webinar-based or Skype-based discussions (cf. [Yarbrough, 2007b](#)). We shared multiple drafts and worked with wiki-pages and other file-sharing mechanisms available at the time to maintain shared focus on comments coming from stakeholders. We developed and shared dozens of drafts to elicit further input and collaboration. To maintain the collaborative process, we directly responded to all stakeholders who sent us suggestions or comments (with the exception of those provided anonymously).

The change from a controlled, top-down, managed process of standards development using pre-existing schema to a more horizontal approach was not always smooth or easy. In spite of ongoing efforts to encourage as much input and feedback as we could, individual stakeholders had to decide when and how to become part of the process. One particularly illustrative controversy concerned the ordering and placement of the utility standards. Early in the development process, some stakeholders had argued that the book should place accuracy, feasibility, and propriety in alphabetical order before utility because all three were viewed as critical prerequisites for greatest utility. We decided that this controversial issue needed to be fully discussed rather than decided without investigation and on the basis of past precedent only. We also presented the draft standards in different orders during national hearings to encourage fresh thinking and to be sure that none of the standards, by virtue of their placement in the order of presentation, systematically received too little attention and review. We fully explained the rationale for presenting the standards in different orders at national hearings in annual written reports to the full JCSEE and also during the national hearings. Nevertheless, these attempts to open up these structures for collaborative deliberation were misunderstood by some previously uninvolved stakeholders who concluded that we members of the task force were intent on changing the order of standards presentation in the final book. Ultimately, however, the allegations and controversies that resulted probably did nothing more than get more stakeholders involved in the process.

This example illustrates that collaborative standards development can be quite political in unexpected ways. It might also illustrate a reaction to the new horizontally focused developmental approach that encourages all stakeholders' ownership and collaboration with regard to all aspects of new editions of program evaluation standards. All who have engaged in collaborative evaluations know that collaborations are never perfect, even when we try hard to make them as effective as possible.

In summary, the approaches described in this article based on recent evaluation scholarship addressing use and collaboration in evaluation seem to have served the third edition standards development well. For the integration of knowledge from this scholarship into the standards development processes we can

thank Lyn Shulha (and other engaged collaborators). Time will tell how successful the third edition is as a product, but in terms of the value of this scholarship to the standards development processes, the evidence and experiences during the development of the third edition strongly recommend it.

ACKNOWLEDGEMENTS

The author thanks Julie Kearney, Valerie Moody Decker, Patricia Martin, and three anonymous reviewers for helpful suggestions on an earlier draft of this manuscript.

REFERENCES

- Alkin, M. C., Braden, J., Frechtling, J., Stevens, F. I., Taut, S., Welner, K., & Vo, A. T. (2008, September). *Validation panel report: Program Evaluation Standards, 3rd Edition*. Retrieved from <http://www.jcsee.org/wp-content/uploads/2010/04/PgES-Validation-Panel-Report-Final-2.pdf>
- Alkin, M. C., & Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1–12. [http://dx.doi.org/10.1016/S0191-491X\(03\)90001-0](http://dx.doi.org/10.1016/S0191-491X(03)90001-0)
- American Evaluation Association Diversity Committee Cultural Reading Task Force. (2004, September). *A cultural reading of the program evaluation standards (2nd edition)* [Final report]. Retrieved from the American Evaluation Association website: <http://www.eval.org/p/cm/ld/fid=74>; <http://www.eval.org/d/do/39>
- American National Standards Institute. (2010, January). *ANSI essential requirements: Due process requirements for American national standards*. Retrieved from <http://www.nist.gov/itl/ansi/upload/2010ANSIEssentialRequirements.pdf>
- American National Standards Institute. (2017, January). *ANSI essential requirements: Due process requirements for American national standards*. Retrieved from https://share.ansi.org/shared%20documents/Standards%20Activities/American%20National%20Standards/Procedures,%20Guides,%20and%20Forms/2017_ANSI_Essential_Requirements.pdf
- Bennett, L. M., & Gadlin, H. (2012). Collaboration and team science: From theory to practice [EB symposium manuscript]. *Journal of Investigative Medicine*, 60(5), 768–775. <http://dx.doi.org/10.2310/JIM.0b013e318250871d>
- Botcheva, L., Shih, J., & Huffman, L. (2009). Emphasizing cultural competence in evaluation. *American Journal of Evaluation*, 30(2), 176–188. <http://dx.doi.org/10.1177/1098214009334363>
- Cousins, J. B. (2004). Commentary: Minimizing evaluation misuse as principled practice. *American Journal of Evaluation*, 25(3), 391–397. <http://dx.doi.org/10.1177/109821400402500311>
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397–418. <http://dx.doi.org/10.3102/01623737014004397>
- Cousins, J. B., & Shulha, L. M. (2006). A comparative analysis of evaluation utilization and its cognate fields of inquiry: Current issues and trends. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation: Policies, programs and practices*

- (pp. 266–291). Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781848608078.n12>
- Cousins, J. B., Shulha, L. M., Whitmore, E., Al Hudib, H., & Gilbert, N. (2016). How do evaluators differentiate successful from less-than-successful experiences with collaborative approaches to evaluation? *Evaluation Review*, 40(1), 3–28; Advance online publication. <http://dx.doi.org/10.1177/0193841X16637950>
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 3–23. <http://dx.doi.org/10.1002/ev.1114>
- Cousins, J. B., Whitmore, E., & Shulha, L. (2013). Arguments for a common set of principles in collaborative inquiry in evaluation. *American Journal of Evaluation*, 34(1), 7–22. <http://dx.doi.org/10.1177/1098214012464037>
- Fetterman, D. W. (2001). *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage.
- Henry, G. T. (2003). Influential evaluations. *American Journal of Evaluation*, 24(4), 515–524. <http://dx.doi.org/10.1177/109821400302400409>
- Hood, S. (2005). Culturally responsive evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 96–100). Thousand Oaks, CA: Sage.
- House, E. R., & Howe, K. R. (2003). Deliberative democratic evaluation [Evaluation theory]. In T. Kellaghan & D. L. Stufflebeam, (Eds.), *Kluwer International Handbooks of Education: Vol. 9. International handbook of educational evaluation* (pp. 79–100). Dordrecht, the Netherlands: Kluwer. http://dx.doi.org/10.1007/978-94-010-0309-4_7
- Joint Committee on Standards for Educational Evaluation (with Stufflebeam, D. L.). (1981). *Standards for evaluations of educational programs, projects, and materials*. New York, NY: McGraw-Hill. (First edition of program evaluation standards)
- Joint Committee on Standards for Educational Evaluation (with Stufflebeam, D. L.). (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (with Sanders, J. R.). (1994). *Program evaluation standards: How to reassess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage. (First edition of program evaluation standards published as *Standards for evaluations of educational programs, projects, and materials*, 1981)
- Joint Committee on Standards for Educational Evaluation. (2003, September). [Minutes of the 29th Annual Meeting]. MiWestern Michigan University, Kalamazoo. Retrieved from the Joint Committee on Standards for Educational Evaluation Minutes website: <http://www.jcsee.org/wp-content/uploads/2009/08/JCMinutes03.PDF>
- Joint Committee on Standards for Educational Evaluation. (2004, September). [Minutes of the 30th Annual Meeting]. National Association of Elementary School Principals, Alexandria, VA. Retrieved from the Joint Committee on Standards for Educational Evaluation Minutes website: <http://www.jcsee.org/wp-content/uploads/2009/08/JCMinutes04.PDF>
- Joint Committee on Standards for Educational Evaluation. (2008, September). [Minutes of the 34th Annual Meeting]. American Psychological Association, Washington, DC. Retrieved from the Joint Committee on Standards for Educational Evaluation Minutes website: <http://www.jcsee.org/wp-content/uploads/2009/08/2008-MINUTES-annual-mtg-FINAL.pdf>

- Joint Committee on Standards for Educational Evaluation (with Gullickson, A. R., & Howard, B. B.). (2009). *The personnel evaluation standards: How to assess systems for evaluating evaluators* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Joint Committee on Standards for Educational Evaluation. (2016). *Program evaluation standards statements*. Retrieved from <http://www.jcsee.org/program-evaluation-standards-statements>
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5–23. <http://dx.doi.org/10.1002/ev.1188>
- Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood, R. K. Hopson, & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory & practice* (pp. 21–39). Greenwich, CT: Information Age.
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57. <http://dx.doi.org/10.1177/1356389004042326>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- O'Sullivan, R. G. (2004). *Practicing evaluation: A collaborative approach*. Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412985468>
- Sanders, J. R. (1994). The process of developing national standards that meet ANSI guidelines. *Journal of Experimental Education*, 63(1), 5–12. <http://dx.doi.org/10.1080/00220973.1994.9943822>
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *American Journal of Evaluation*, 18(1), 195–208. <http://dx.doi.org/10.1177/109821409701800121>
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation*, 37(2), 193–215. <http://dx.doi.org/10.1177/1098214015615230>
- Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4(3–4), 181–201. [http://dx.doi.org/10.1016/0004-3702\(73\)90011-8](http://dx.doi.org/10.1016/0004-3702(73)90011-8)
- Thomson, A. M., Perry, J. L., & Miller, T. K. (2007). Conceptualizing and measuring collaboration. *Journal of Public Administration: Research and Theory*, 19(1), 23–56. <http://dx.doi.org/10.1093/jopart/mum036>
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wood, D. J., & Gray, B. (1991). Toward a comprehensive theory of collaboration. *Journal of Applied Behavioral Science*, 27(2), 139–162. <http://dx.doi.org/10.1177/0021886391272001>
- Yarbrough, D. B. (2005a). *Program evaluation standards development task force annual report for 2005*. Iowa City, IA: University of Iowa, Center for Evaluation and Assessment.
- Yarbrough, D. B. (2005b). *Report of the task force survey results and evaluation of the program evaluation standards, 2nd Edition*. Paper presented at Evaluation 2005, the second international conference on evaluation sponsored by the American Evaluation Association and the Canadian Evaluation Society, Toronto, ON.

- Yarbrough, D. (2007a). *Key issues regarding revisions in format, content, and organization of the PgES3*. Presentation at the American Educational Research Association Annual Meeting, Chicago, IL.
- Yarbrough, D. B. (2007b). *Program evaluation standards development task force annual report for 2007*. Iowa City, IA: University of Iowa, Center for Evaluation and Assessment.
- Yarbrough, D. B. (2016). *Table of utility quality standards and process criteria and sub-criteria in evaluations, revised April 2016*. Iowa City, IA: University of Iowa, Center for Evaluation and Assessment.
- Yarbrough, D. B., Shulha, L. M., & Caruthers, F. (2004). Background and history of the Joint Committee's program evaluation standards. *New Directions for Evaluation*, 2004(104), 15–30. <http://dx.doi.org/10.1002/ev.133>
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.
- Yarbrough, D. B., Yarbrough, N. D., Karatoprak, R., Liu, W., Tibbetts, K., & Waiyavutti, C. (2016). *Challenges and issues in interpreting and applying the North American program evaluation standards in multiple cultural locations*. Paper presented at the 12th Biennial Conference of the European Evaluation Society, Maastricht, the Netherlands.

AUTHOR INFORMATION

Donald B. Yarbrough, Ph.D., is Professor of Educational Measurement and Statistics and Professor of Educational Policy and Leadership Studies at the University of Iowa, USA. He was the founding director of the University of Iowa Center for Evaluation and Assessment (1992–2015), Chair of the Joint Committee on Standards for Educational Evaluation (2008–2014, member 1996–2014), and Chair of the Program Evaluation Standards Development Task Force (2003–2010). He is a practicing program evaluator and has conducted with collaborators more than 150 funded evaluations. From 2008 to 2014, he was also the associate director for evaluation and assessment in the Institute for Clinical and Translational Science, University of Iowa Carver College of Medicine. His current funded research focuses on STEM and health-related education and evaluation.